

On Novel System for Detection Video Impairments Using Unsupervised Machine Learning Anomaly Detection Technique

Nermin Goran¹, Alen Begović¹, Alem Čolaković¹

¹ *Univesity of Sarajevo, Zmaja od Bosne, Sarajevo, Bosna and Herzegovina*

Abstract – Recently, the necessity of video testing at the point of reception has become a challenge for video distributors. This paper presents a new system framework for managing the quality of video degradation detection. The system is based on objective video quality assessment metrics and unsupervised machine learning techniques that use the dimensionality reduction of time series. It was demonstrated that it is possible to detect anomalies in the video during video streaming in soft real time. In addition, the model discovers degradations based on the visible correlation between adjacent images in the video sequence regardless the quick or slow change of a scene in the sequence. With additional hardware manipulations on the equipment on the user side, the proposed solution can be used in practical implementations where the need for monitoring possible degradations during video streaming exists.

Keywords – Anomaly detection in video sequence, IPTV, QMS, SSA analysis, unsupervised learning model, video impairments.

1. Introduction

The ubiquity of video services problematizes two important research issues in this area.

DOI: 10.18421/TEM124-10

<https://doi.org/10.18421/TEM124-10>

Corresponding author: Nermin Goran,
*Univesity of Sarajevo, Zmaja od Bosne, Sarajevo,
Bosna and Herzegovina.*


Email: nermin.goran@fsk.unsa.ba

Received: 03 July 2023.

Revised: 13 September 2023.

Accepted: 05 October 2023.

Published: 27 November 2023.

 © 2023 Nermin Goran, Alen Begović & Alem Čolaković; published by UIKTEN. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

The article is published with Open Access at <https://www.temjournal.com/>

The first one is finding new compression techniques so that video can be more effective to store and/or end-to-end transmitted through the telecommunications system [1], and the second one is how this process affects the Human Visual System (HVS) [2]. Considering this from the perspective of telecommunication network, Quality of Service (QoS) and consequently Quality of Experience (QoE), i.e., fulfilment of user expectation with the service, are the main issues that have to be addressed [3]. In era of 5G and FTTH networks, we encounter the decrease of potential QoS problems which are introduced as latency, jitter, and packet loss in access and radio domain, regarding the higher service rate provided by wider bandwidth [4]. At the same time, with using Software Defined Networks (SDN) QoS aware or legacy Multiprotocol Label Switching (MPLS) networks, QoS issues in network layer should be resolved properly. However, increase of bandwidth will not solve QoS problems in heterogeneous and legacy access fixed or wireless networks. In these environments, there are many other parameters which should be controllable in order to achieve proper and good quality of video service like appearance of fading, interference, and noises. If this requirement is not fulfilled, various types of video degradation can still occur. [5]. For example, different kinds of fading in time variant wireless communication channel or appearance of nonstationary noise in DSL can generate a huge packet loss and other QoS issues that can cause poor video quality and consequently low user satisfaction, although there is sufficient bandwidth [6]. Thus, understanding the relationship between appearance of QoS problems and user perceived quality of service is an important step in creating good Quality Management System (QMS) [7]. Beside this, QMS needs to recognize a degree of unsatisfactory effects which should alert service provider about appearance of a potential problem without user involvement. Considering multimedia quality analysis, packet loss can cause a loss of video macroblocks of a frame or even loss of multiple frames which depends on the size of packet loss [8], [9].

Also, delay and jitter have a big impact on audio quality and synchronization as well as video reproduction. All this leads to a poor user satisfaction with the service especially in case when video was degraded, because audio generates less disturbance effects. It shows that QMS is needed for monitoring of good or bad states.

Quality of video service can be evaluated using objective and subjective methods [10]. Subjective methods are based on user evaluations and they request sufficient number of human evaluators which is very hard to collect especially if real time video services are taken into consideration. New crowdsourcing techniques are promising about easier and better subjective testing [11]. On the other hand, objective methods are cheaper and easy implemented, but most of them have a very weak correlation with subjective methods. Currently, there are three accepted models for objective quality evaluations: full reference (FR), reduced reference (RR) and no reference (NR) models [12]. In case of FR, original and tested sequences have to be present at the user side during quality analysis. For the RR method evaluation, the parts of video signal or some video information should be distributed at the user's side. It's obvious that using these methods during real time video, and for so needed quality evaluation, we should transfer or have additional video or additional video information. Consequently, the both mentioned methods are not very suitable as practical solutions. Considering this, the most elegant and the least demanding method but mathematically demanding is NR method which can use techniques of mathematical analysis or consistency and properties in video signal in order to satisfy proper criteria of video quality testing. Recently, there are lots of promising NR models which use machine learning or deep learning techniques for better video quality analysis [13], [14]. Most of objective metrics for video quality assessment are based on image-by-image (frame by frame) quality analysis. There are lots of techniques which are recommended for testing a video sequence such as Structural Similarities (SSIM), Video Quality Metrics (VQM) or Mean Squared Error/Peak Signal to Noise Ratio (MSE/PSNR) [15]. The main focus for those techniques is structural image analysis or MSE of image pixels. Recently, some authors worked on including motion during video testing in order to conduct better in video analysis NETFLIX [16] and V-SSIM [17]. The focus of authors is on the problem of NR IQA for blurred images and they propose a new No-Reference Structural Similarity (NSSIM) metric based on re-blur theory and (SSIM) [18].

Although many image-based quality analysis techniques have a weak correlation with subjective techniques we can easily use them at least for discovering video degradations especially at user's side. In many cases decoding a video sequence leads to selection of reference frame form Group of Pictures (GoP) and appropriate motion vectors that are needed for a proper video stream. Video degradations, caused by packet loss or other violations of QoS indicators, can appear in some frames or parts of the frame, and along with decoding process, they lead to worst user satisfaction.

It is very important to emphasize, when packet loss happens, these degradations very often affect more adjacent frames of a video sequence. In addition, in this case, there are certain correlation phenomena that appear and can be observed during this occurrence. Regarding this, we developed the model which can discover those video degradations and raise alarm in order to inform service provider that something is happening at user's perspective without user involvement in testing process. This is very important especially in real time video distribution such as Internet Protocol television (IPTV). IPTV service is a platform which uses MPEGTS/UDP transport protocols with constant datagram stream. In that case, there are no alarm discovery procedures which can identify all losses or other corresponded problems at the side of a video consumer. This is left to the lower OSI layers (data and physical) which are not so effective and can make additional errors. Based on all mentioned facts, we can conclude that the solution which could discover the moment when users have a problem with proper video streaming is necessary for building a good quality management system QMS especially for service and network providers. In these situations, the video enhancement system described in [17], [19] would also be useful for this purpose. However, our goal is to explain the QMS system for detecting these phenomena.

The creation of our model primarily addresses this core issue. In the second section, we introduced framework of potential QMS that can use mathematical model for video quality detection. The model, proposed in this paper, finds parts of video sequence with structural degradations or noise by observing adjacent images of video sequence at customer's side. It is based on presumption that there is high correlation between adjacent decoded images which create a video sequence, and violation of the correlation is a reliable notification for existence of degradations.

For image comparing and discovering correlation between them, we can use any of the objective methods and metrics. In the third section we deal with usefulness of SSIM and other known objective metrics. In our model we chose modified SSIM metrics based on structural correlations. We used calculated SSIM values between adjacent images and made time series that is examined to evaluate the entire video sequence over a period of time. This time period was a period which covered the duration of a video sequence in arbitrary number of images per seconds. After collecting this data, the model used unsupervised machine learning with anomaly detection techniques which discovered changes in time series that depended on video changes. These changes between adjacent images could be smaller or bigger which depended on the size of degradation and structure distortions between images during the video reproduction. In the fourth section we explained how model works and we discuss mathematical interpretations of the model. In these sections, we consider the correlation, singular values and variances of time series of the assessed video sequences using mathematical analysis to find distortions between adjacent images/frames. The main aim is building useful QMS for checking proper video streaming at the end user. In this way, one can easily estimate the quality of the video that is delivered over lossy networks without sending technicians on the field or to the user. In the fifth section test values and results are presented, following closing conclusion.

2. QMS Framework of Video Detection

Figure 1 shows a new framework for detection of video distortions or received video quality at the user's side. We supposed, based on our practical experience in service and network maintenance of IPTV network, that receiver (Set Top Box (STB)) is the best point for video quality analysis. If the problem that caused video distortions occurs anywhere in the network shown in Figure 1, it will be noticed at the STB's outputs.

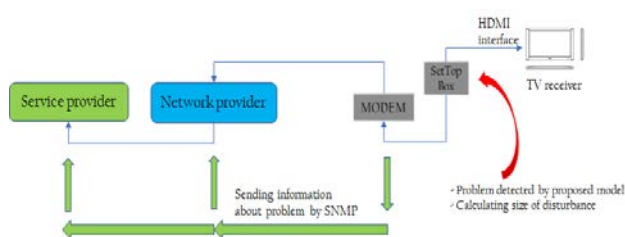


Figure 1. Framework for video quality detection

In IPTV applications, the common situation is that after video distribution from service provider headend over telecommunication network of network provider, video decoding is happening at STB and TV receives YUV images over High-Definition Multimedia Interface or HDMI interface. We proposed that these images could be collected for the purpose of video quality testing on image-by-image basis. After image collection, we could generate software script which could be installed into TV receiver or STB. That piece of software would have checked the state of video distortions in a received video sequence that consisted of particular images using unsupervised anomaly detection method. When video distortions would be discovered by the script (problem detected) then the size of video degradation could be calculated and service or network provider could be informed about these conditions. Image quality can be checked by some known methods for objective quality analysis like SSIM. It should be emphasized without neglecting generalizations, beside SSIM, it is possible to use any image-by-image basis quality metrics (MS-SSIM, VQM, PSNR or Video Multi-Method Assessment Fusion (VMAF)) for finding degradations in a video sequence. In our case, we used only a part of SSIM metric which is based on correlation between two adjacent images. Obtained correlation values can be used for creation of time series for analysis, described in section 4. For comparison, the MSE is used in order to prove functionality of the model with another technique.

3. Objective Metric of Quality

SSIM method compares structural similarity (correlations) between images contrary to traditional approaches such as MSE and PSNR, which are typically implemented to detect real distortion in a video, obtained after compression or from error-prone transmission channels [20]. Original images which are compared using the MSE and PSNR metrics, have a very low correlation to HVS. In order to find precise evaluation and better correlation with HVS the authors [21] proposed mentioned SSIM method. SSIM method is designed to ignore changes caused by illumination; contrast and mean values of the image. When two images are compared using SSIM, they are compared considering the smallest structural image blocks. Then, if there are differences between image blocks, SSIM method results in small values (near 0) and contrary, when two image blocks are similar, SSIM values are nearer to 1. Therefore, when video distortions or differences between images happen, SSIM method works best.

Considering appearance of video degradations such as blocking, blurring, noise, or other compression artefacts.

To design an SSIM model, it is necessary to measure three parameters separately: the luminance, contrast, and structure of the image.

Suppose we compare two images (or a small block of images) X and Y, mean values of each image μ_x and μ_y are:

$$\mu_x = \frac{1}{N} \sum_{n=1}^N x_n; \mu_y = \frac{1}{N} \sum_{n=1}^N y_n \quad (1)$$

where x_n and y_n are pixels of each image, respectively. Beside this, we have to assume that distribution of pixels is uniform.

The standard deviations of each image (s_x and s_y) can be calculated by:

$$s_x = \sqrt{\frac{1}{N} \sum_{n=1}^N (x_n - \mu_x)^2}; \quad (2)$$

$$s_y = \sqrt{\frac{1}{N} \sum_{n=1}^N (y_n - \mu_y)^2}$$

where we assumed that there is a significant number of pixels. Considering (2), correlation between two images (X and Y) equals:

$$\rho_{xy} = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y} \quad (3)$$

$$= \frac{\sum_{m=1}^N \sum_{n=1}^N (x_m - \mu_x) \cdot (y_n - \mu_y)}{\sqrt{\sum_{m=1}^N (x_m - \mu_x)^2 \sum_{n=1}^N (y_n - \mu_y)^2}}$$

where σ_x and σ_y are variances of two standard deviations s_x and s_y , respectively. We should calculate values from (1) to (3) to obtain unique SSIM value. Therefore, regarding [21], SSIM consists of three parts:

$$\text{SSIM}(x,y) = l(x,y)^\alpha \cdot c(x,y)^\beta \cdot \rho_{xy}^\gamma \quad (4)$$

The first part of (4) represents luminance ($l(x,y)^\alpha$), the second contrast ($c(x,y)^\beta$) and the third structural similarities or correlations (ρ_{xy}) between images or parts of images. All these parts are combination of μ_x , μ_y , σ_x , σ_y , ρ_{xy} , where α, β, γ are empirical constants and their mathematical calculations can be found in [19].

The weakness of SSIM is the insensitivity to distortions in magnitude of intensity in relation to distortions resulting from spatial displacement distortion. Human HVS scores more on video quality in the case of spatial displacement distortion. Another weakness is found in the comparison between blurring and white noise distortion. It is shown that even though the presence of white noise in the image is obvious, the SSIM values are too low for this type of distortion. This indicates that SSIM is insensitive to changing the brightness of the image. SSIM has been shown to have an advantage over traditional approaches, but it also has limitations in important areas of image distortion. Although there are promising evaluation models [22], we used partially widely accepted SSIM in order to find inconsistency in images or precisely in order to find distortions between adjacent images in a video sequence. The proposed model in the next section can use any metrics which can measure the correlation between two regions inside of images or between two whole images. We used only the correlation that is a part of the SSIM metric ($\rho(x,y)$ part in (4)) only to check the changes in adjacent images in a simple way. In continuation of this section, other metrics for quality evaluations that can be used in the same manner will in short be explained. The first of them is MSE (or PSNR) [20]. MSE is the metric which is based on finding differences between the luminance values of correspondent block of pixels in original and tested images. MSE equals:

$$\text{MSE} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M ([x(i,j) - x'(i,j)])^2 \quad (5)$$

where $x(i,j)$ and $x'(i,j)$ are pixel values of N or M pixels inside two images respectively. Using MSE the PSNR can be found as:

$$\text{PSNR} = 10 \log_{10} \frac{\text{MAX}^2}{\text{MSE}} \quad (6)$$

where MAX is maximum pixel value of the whole reference image. The next two objective techniques are MS-SSIM [23] and VSSIM [17] which are based on SSIM metrics. The first one uses down-sampling of original images and calculates the difference between image details at different resolutions. The second combines motions and SSIM in order to calculate a unique video quality measure. VQM (Video Quality Metric) explained in [24] is the model which is based on DCT (Discrete Cosine Transform) analysis of original and tested video frames.

This model conducts DCT for 8x8 matrixes of pixels of original and tested image in order to find the contrast for LC_o and LC_t values of each image:

$$LC_{o/t} = \frac{DCT(i, j) \cdot (\frac{DC}{1024})^{0.65}}{DC} \quad (7)$$

After this process LC_o and LC_t are converted in just noticeable difference values using static and dynamic spatial contrast sensitivity function. These values are subtracted in order to get unique Diff(t) value and VQM score is:

$$VQM = Dist_{mean} + 0.005 \cdot Dist_{max} \quad (8)$$

where:

$$Dist_{mean} = 1000 \text{ mean}(\text{mean}(\text{Diff}(t))) \quad (9)$$

and

$$Dist_{max} = 1000 \text{ max}(\text{max}(\text{Diff}(t))) \quad (10)$$

Another promising technique is Video Multimethod Assessment Fusion (VMAF) explained in [16]. It is NETFLIX modern technique for video quality testing based on information fidelity loss at four different spatial scales and impairments which distract viewer attention which has good correlation with different subjective video datasets. Considering all above metrics, we will neglect precision of different measures because the main aim of our quality detection model is finding degradations in images but not comparison of techniques. For our model any of these techniques can be used yet the one with correlations between images (SSIM) is chosen because it has the best fit with the model. For comparison MSE is also used.

4. The Model Description of Extension of Objective Metrics

To define and describe the model, it was necessary to analyze a potential system that can recognize image quality and detect potential anomalies in the image. In addition, this section will also present mathematical equations that show how the model should work and how it helps the system to detect degradations in the video. By using them, it is also possible to determine the size of the degradation.

4.1. Model Description

It is known that each video consists of images and has a certain frames rate per seconds.

In addition, video can have different I, P or B frames. These frames create GoP (Group of Pictures) which is the basic element for video decoding process. In order to achieve “time-part” of video compression the coding process uses GoP to make a reference between the previous and the next frames in order to estimate proper video motion vectors.

Beside GoP, the reference can also be an object or a part of previous or next frames especially as the novel coding technique. Whatever coding video process is, the result of decoding is a group of YUV images which feeds input or HDMI interface of a TV receiver. When we are evaluating the quality of two videos, in fact we need to compare all decoded images that video consists of. It should not be neglected that, during decoding video process, a lot of images generate in accordance with a known parameter that is called a number of frames per seconds (fps).

Considering this, there are a lot of models of objective video quality evaluation which are based on separation and testing a huge number of images which create a video sequence to estimate QoE in offline regime [17], [25]. The main aim in this particular evaluation of every image is obtaining the mean value of video quality. Thus, every evaluation can be plotted on time graph where x axis indicates a number of images and y axis indices a value of testing of a particular video image. Therefore, the first condition for establishing a model is the analysis of the video as a series of images and the assessment of the quality of each adjacent video image, which contributes to the possibility of creating our model. We noticed a very important assumption which appears during the testing and comparing of adjacent images of a video stream. Adjacent images of video stream have a high degree of correlation regardless of rate of scene changes and that was mentioned in study [10]. In addition, we concluded that there is a low correlation between adjacent images with additional distortions (artefacts) and that correlation can be low or lower from case to case. These assumptions are used for creating our model which can discover images with visual distortions as well as the size of the distortions. We also proposed that mutual structural comparisons of every video frame in a video sequence can be represented as a time series. Accordingly, the time series is formed of evaluated values which represent correlation between adjacent images. Process of monitoring these values implies finding degradations as anomalies in time series i.e. series of images. Therefore, the model discovers appearance of structural or other changes in the quality of particular images and thus the quality of whole video sequence or video sequence of interest.

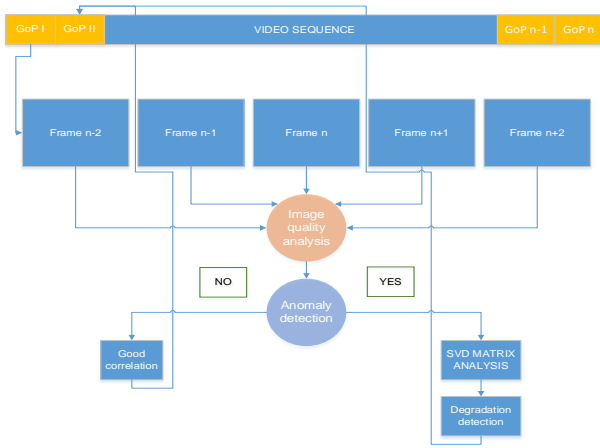


Figure 2. Model description

Figure 2 depicts a simple pictorial explanation of algorithm which is used for anomaly detection in a video sequence.

After the video sequence is decoded and partitioned in a group of pictures (unbonded with real GoP structure) the model conducts a comparison inside of groups with a common method for objective quality image testing.

We used SSIM and MSE, because of their correlation basis, although it is not excluded to use any other techniques for objective quality evaluation. The result of the first step is a series of values which represent correlation between adjacent images. The anomaly detection block conducts SSA analysis of series which is divided in small matrices.

If there is no anomaly in matrices, a good correlation between images happens and the model can grab the next GoP. In the contrary, if there is an anomaly in series, the next block should discover the size of visual degradation inside the image and the number of affected images i.e., its duration. For this purpose, the model uses SVD analysis, trace of singular matrices and their means and variances. In this way, we discover how many degradations frames it covers with what degree of damage.

4.2. Mathematical Description

As mentioned in the previous section, a video sequence consists of images which flow with specified rate considering that we divided the video on specified number of images per second before analysis and after video decoding. In this case we have a specified number of images that is independent of the codec used previously. After this we conducted a correlation check using the equation (1-3) between every adjacent image and formed time series of these values. This time series can be represented as vector $\rho_{x_1y_1}, \rho_{x_2y_2}, \dots, \rho_{x_ky_k}, \dots, \rho_{x_ny_n}$ where n represents a number of mutual correlations between video images (frames).

SSA analysis is the first stage of anomaly detection. The first step of SSA analysis according to [26] is embedding process. Embedding is forming of Hankel matrix of original lagged time $\rho_{x_1y_1}, \rho_{x_2y_2}, \dots, \rho_{x_ky_k}, \dots, \rho_{x_ny_n}$ and it equals:

$$H = \begin{bmatrix} \rho_{x_1y_1} & \rho_{x_2y_2} & \dots & \rho_{x_{n-k}y_{n-k}} \\ \rho_{x_2y_2} & \rho_{x_3y_3} & \dots & \rho_{x_{n-k-1}y_{n-k-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{x_ky_k} & \dots & \dots & \rho_{x_ny_n} \end{bmatrix} \quad (11)$$

where k represents a number of lagged sequences. Autocorrelation matrix of H is:

$$R = \frac{1}{n} HH^T \quad (12)$$

and it represents the outer product of each column vector from matrix H. Singular value decomposition of autocorrelation matrix R can be conducted with:

$$SVD(R) = U \Sigma V^T \quad (13)$$

where U and V are matrix of singular vectors of R and Σ is the matrix of singular values of R. In this case, because R is symmetric matrix, vectors V^T and U are orthogonal and equal and so we can write $SVD(R) = U \Lambda U^T$, where Λ is the matrix of eigenvalues. On the other hand, according to [27], it's possible to take a piece of original time series in order to discover anomalies in the observed series. In our case, we divided matrix H in i small matrixes h_i of less dimensionality which can be used for so called "windowing" of the series. It implies the reduction of matrix dimensionality.

As we mentioned in the introduction, we suppose that there is a high correlation between adjacent images which are not degraded so that these matrixes can be used to track the correlation of smaller group of images. For each reduced matrix we can also find autocorrelation matrix:

$$r_1 = \frac{1}{i} h_i h_i^T \quad (14)$$

and conduct singular value decomposition of reduced autocorrelation matrix:

$$SVD(r) = u \xi_l u^T \quad (15)$$

The (15) represents two matrixes of eigenvectors (u, u^T) and one matrix of eigenvalues ξ_l . For our further examination the matrix of eigenvalues ξ_l was the crucial. According to [28] a number of nonzero eigenvalues defines rank matrix r_l . In case of high correlation, matrix $\xi_l = [\sigma_1, 0, \dots, 0]$ has singular values where the first singular value will be the highest value while others are zero or close to zero.

On the other hand, we can calculate the trace of matrix $Tr(\xi_l) = \sigma_1 + \sigma_2, \dots, \sigma_{lg}$ where lg is a number of lags in matrix r_l i.e., the size of window of small matrix h_i . In case of good correlation, we can neglect other eigenvalues and the trace of a matrix ξ_l will be $Tr(\xi_l) \approx \sigma_1$. Furthermore, considering these characteristics if there is degradation in image, we will notice a low correlation between adjacent images and consequently we can conclude that degradations happen. Therefore, if there is a degradation in time series then the number of nonzero elements of singular matrix ξ_{l+1} increases in comparison with the previous matrix ξ_l . The most important consequence of this state is the fact that the correlation within the observed pieces of series which are collected in matrix r_{l+1} is lesser than in the previous series (values of matrix r_l). In these cases, first eigenvalue in the previous matrix ξ_l is higher than the first eigenvalue in the next matrix ξ_{l+1} i.e., $\sigma_{l_1} > \sigma_{l+1_1}$, but other eigenvalues will be higher in the next matrix in comparison to eigenvalues of the previous matrix $[\sigma_{l+1_m} > \sigma_{l_m}, \text{ where } m = 2, 3, \dots, lg]$. In addition, appearance of additional nonzero singular values in the observed (next) matrix shows that there is degradation in time series. The conclusion is that if we consider the video sequence as a time flow of images and form time series of the flow, we can discover the degradation of video sequence by observing matrix representation of correlations between particular images. In the same time, the trace of matrix can discover the size of degradation in the observed matrix considering number of non-singular elements. According to [29] the trace of singular matrix represents energy distribution of source matrix or in this case time series. In the case of good correlation, the power will be concentrated in the first singular values while in the case of degradation the power of singular values in matrix will be distributed equally especially in the case of high disturbance.

In addition, as we mentioned, the trace of the previous matrix is higher than the next matrix $Tr(\xi_l) > Tr(\xi_{l+1})$ and singular values are distributed according to this relation $\sigma_{\xi_{l_1}} > \sigma_{\xi_{l+1_1}}$, but $\sigma_{\xi_{l_2}} < \sigma_{\xi_{l+1_2}}, \dots, \sigma_{\xi_{l_l}} < \sigma_{\xi_{l+1_l}}$. Beside this distribution, the sum of previous and observed matrix $sum(Tr(\xi_l)) > sum(Tr(\xi_{l+1}))$ and variance $var(Tr(\xi_l)) > var(Tr(\xi_{l+1}))$ can be a good indicator that there is a huge or small degradation. On the contrary, when the observed matrix reverts to the previous one, and a new observed matrix emerges with improved correlation, the same comparisons but with inverted inequalities can be applied, leading to distinct conclusions as a result. In that case, the process continues as in the beginning i.e., assuming a high correlation between images.

To summarize, if we observe a time series through the described mathematical relation (11-15) and described process then it is possible to discover the appearance and the degree of degradation in a video sequence.

5. Testing and Results-Testbed Descriptions

Testing was conducted in real IPTV network at a network provider in Bosnia and Herzegovina. Original video with sport content was delivered over IPTV network in H.264 format and recorded by VLC (Video LAN Client) on the hard disk at the user side. We recorded a 40-second video in .ts format, containing a total of 1000 frames (images), as it complied with the standard frame rate of 25 frames per second (fps). It should be emphasized that, after the decoding process, the video was saved and divided into YUV images. We divided images into three channels and for the sake of simplification, we used only grayscale images for further processing. Then, we used objective SSIM method for image quality estimation which consists of unified tree metrics, testing of luminance, contrast, and correlation between two adjacent images. For correlation testing, SSIM gives the unique evaluation ranging from 0 for completely uncorrelated images to 1 for the completely equal images.

After this process, we ran the python script, calculated correlations made database for further anomaly detection. With other python script we created 4x4 lagged matrix from the database (we chose 4x4 but we could take less or high dimensions), divided on small matrices and conducted all mathematical equations. Most of images had high correlations but anomalies discovered between image 6 and 16 were in accordance with Figure 3.

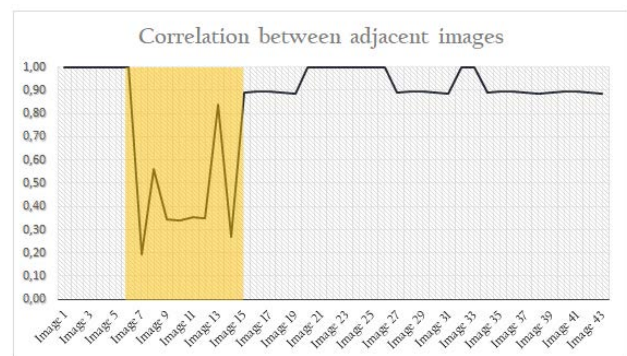


Figure 3. Correlation $\rho_{x_i y_i}$ between adjacent images

Figure 3 depicts a snapshot of correlation values for different images in a row. A shadowed detail with small values, which can be noticed in the Figure 3, shows situations where degradations appear.

We conducted an additional analysis at this point considering equations from mathematical descriptions.

The shadowed part of the figure also shows moments where there is a low correlation between adjacent images when QoS decreased in the network, probably because of packet loss.

These are moments where a structural error appears in images which the model recognizes as anomalies.

Once the anomalies are detected the model should calculate the size of the degradation of the affected images. This was done and explained with mean, sum and variances of singular values. All mathematical calculations from chapter 4.2 were performed using a Python script and is presented in Table 1.

Table 1. SVD, SUM, MEAN and VARIANCE of eigenvalues values

Images	Singular (sin.) values	Sum of sin. values	Mean of sin. values)	Variance of sin. values
1-5	[3.55, 0.12]	3.679	1.839	2.948
6-8	[1.42, 0.11]	1.537	0.769	0.4308
9-11	[0.64, 0.029]	0.671	0.336	0.0948
12-14	[2.25, 0.046]	2.295	1.147	1.214
15-20	[3.15, 4.97e-04]	3.152	1.576	2.482
20-25	[3.58, 2.25e-03]	3.579	1.7897	3.1950
26-43	[3.43, 3.26e-03]	3.498	1.8011	3.0564

5.1. Analysis of the Results

The second column of Table 1 contains the first two singular values of diagonal matrices obtained with SVD analysis of parts of lagged Hankel matrix H. In this particular case (Figure 3), we can notice using assumption from mathematical equations, that the first singular value is very high when there is a high correlation, while in the case of a low correlation there is a decrease of that value.

It can also be noticed that the singular values during the first degradation, which involves three images (6-8), are slightly higher than the singular values related to the degradation in the next group of images (9-11). But, at the same time, these singular values are lower than the values related to the group of images of 12-14. Since after first 15 images there are no further degradations, the adjacent images from 14 to 43 have the high correlation and high the first singular value can be noticed again.

After image 43, there were no significant degradations and consequently high changes in the correlation between adjacent images. Hence, the last singular values in the Table 1 represent a larger range i.e., the singular values were very similar for images from 44 to 1000 (we captured 1000 images in total). In addition, the size of degradation is estimated with sum, mean, and variance. Low mean and low variance show that there are no differences between degraded images and that condition lasts until a better correlation appears. The table excludes certain peaks that are found at each transition of the degraded image to the normal state, which can be observed each time on this occasion. Detection of this condition was determined at a sharply low correlation between the degraded image and the image that is without degradation.

5.2. Pictorial Explanation of Generated Degradation

We used objective SSIM measure (mathematical equation (4)) to calculate correlation between adjacent images. Figure 4 (from a to f) depicts low and high SSIM values or low and high correlation which depend on similarities between images. In the case of a huge degradation there is a low similarity between adjacent images in contrary to good similarity which was noticed in the case without any disturbances in the adjacent images.

The (a) shows situation where there are no degradations. When the degradation happens, SSIM of degraded images (b, c, d) has low values. SSIM of non-degraded images (e and f), as a part of a video sequence when degradation stopped, has a high value again. In the other calculations SSIM was high values, but because of huge video sequence here is depicted just a few of interesting calculations during video degradations. It should be emphasized that all figures and analysis are given using Python script which you can find on GitHub [30].

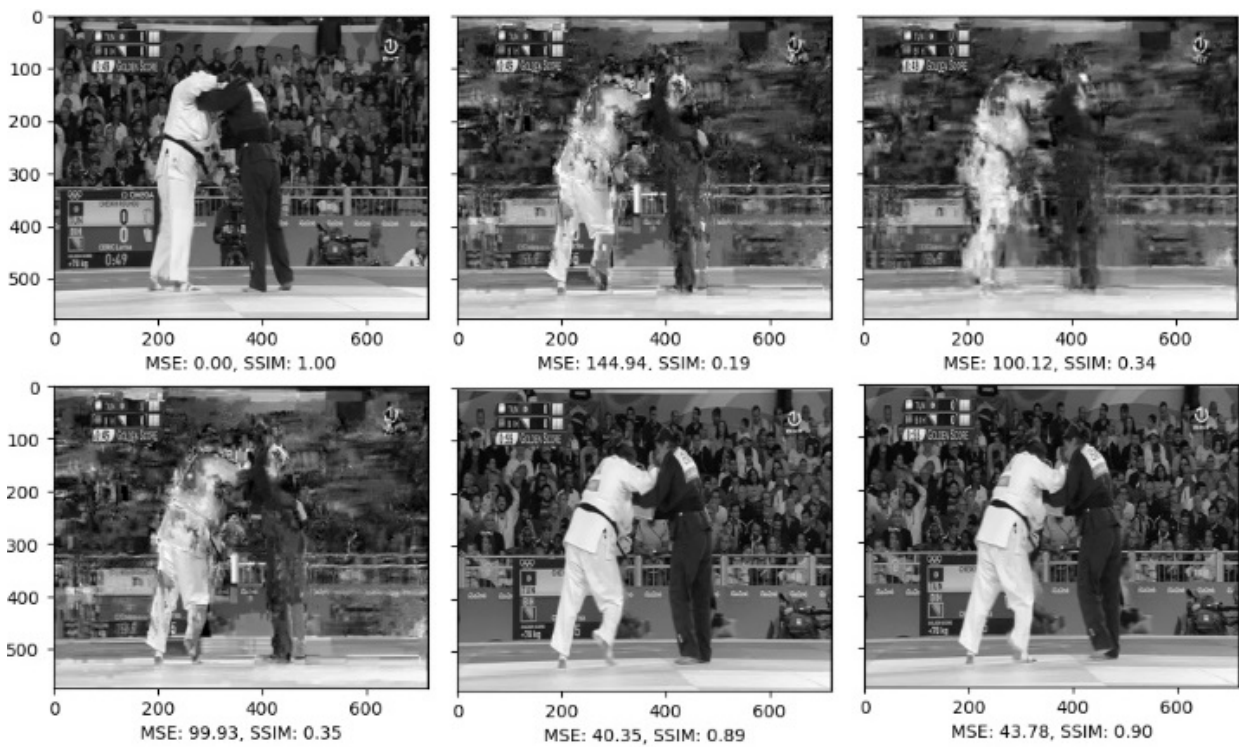


Figure 4. SSIM (correlation) between images a) 4 with 5 images b) 5 with 6 images c) 6 with 7 images d) 7 with 8 images e) 15 with 16 images f) 16 with 17 images

6. Conclusion

Quality of analysed model and conclusions can be summarized in a few theoretical and practical contributions. In this paper we created the model for finding degradations in a video sequence after its decoding at the user side. In this case using any decoding schema in video distribution process is not necessary to investigate. Images after being decoded from their GoP structure with decoding schema for example H.264 or H.265, join in a video sequence with defined frame rate. These images using YUV format go to analogue or using HDMI to digital interfaces of TV receiver which interprets data using EIA/CEA-861 standard as a media stream. This is the point where the solution, described in the paper, can be implemented. Appearance of degradation is a very common at network providers when IPTV service is distributed over MPEG TS and UDP protocol without controlling QoS mechanism in the last mile including home user's network. In these cases, user's QoE decreases in accordance with the decline of QoS. This model using only a few equations, but with very huge computational needs, can discover those cases. The advantage of the model is very simple implementation as it can be noticed considering a simple example of testing video sequence after we captured it on the customer side. Further, implementing of a small piece of hardware in Set-Top-Box with using simple script we can build simple QMS and for example using SNMP messages improve QoS/QoE monitoring.

With the similar form, equations depicted in the model can be useful in other implementations where some anomalies should be detected. Shortcoming of model is need of huge computational power and enhancement of buffers.

For the future works we plan to improve the estimation of degradations considering regions of the image and solve situations with high peaks when degradations disappear. We also plan to use TCP besides UDP in order to check image degradations in this case. We assume that in this case we will have many good correlated images during the video freezing. This state is possible to discover using SSIM metrics because correlation is never 1 but close to 1. In addition, in the future works we plan to analyse all channels (YUV or RGB) and find degradations in all of them.

Acknowledgements

This work was supported by the Federal Ministry of Education and Science, Federation of Bosnia and Herzegovina, Bosnia and Herzegovina.

References:

- [1]. Kumar, K., Kumar, R., & Pandit, A. K. (2019). Effect of Multiple Constraints on Multimedia Data Transmission in High Efficiency Video Coder (HEVC). *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, India, 269-272. Doi: 10.1109/COMITCon.2019.8862229

- [2]. Chen, P., Li, L., Huang, Y., Tan, F., & Chen, W. (2019). QoE Evaluation for Live Broadcasting Video. *2019 IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan, 454-458. Doi: 10.1109/ICIP.2019.8802978
- [3]. Akhtar, Z., Siddique, K., Rattani, A., Lutfi, S. L., & Falk, T. H. (2019). Why is Multimedia Quality of Experience Assessment a Challenging Problem? *IEEE Access*, 7, 117897-117915. Doi: 10.1109/ACCESS.2019.2936470
- [4]. SG12/3GPP. (2022). *Y.IMT-2020.qos-mon*. Geneva: ITU.
- [5]. Ibrahim, I., & Khamiss, N. (2019). Proposed of the wireless mobile system and video coding system in the heterogeneous network. *Multimedia Tools and Applications*, 78(23), 34193–34205. Doi: 10.1007/s11042-019-08230-8
- [6]. Goran, N., & Hadžialić, M. (2017). Mathematical Bottom-to-Up Approach in Video Quality Estimation Based on PHY and MAC Parameters. *IEEE Access*, 5, 25657-25670. Doi: 10.1109/ACCESS.2017.2772042
- [7]. Barakabitze, A. A. (2019). QoE management of multimedia streaming services in future networks: a tutorial and survey. *IEEE Communications Surveys & Tutorials*, 22(1), 526-565. Doi: 10.1109/COMST.2019.2958784
- [8]. Korhonen, J. (2018). Learning-based Prediction of Packet Loss Artifact Visibility in Networked Video. *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, Cagliari, Italy, 1-6. Doi: 10.1109/QoMEX.2018.8463394
- [9]. Torres Vega, M., Perra, C., & Liotta, A. (2018). Resilience of Video Streaming Services to Network Impairments. *IEEE Transactions on Broadcasting*, 64(2), 220-234. Doi:10.1109/TBC.2017.2781125
- [10]. Goran, N., Begović, A., & Škaljo, N. (2020). Adjacent Image Correlation for Video Quality Assessment. *2020 International Symposium ELMAR*, Zadar, Croatia, 45-48. Doi: 10.1109/ELMAR49956.2020.9219047
- [11]. Gardlo, B., Ries, M., & Hossfeld, T. (2012). Impact of screening technique on crowdsourcing QoE assessments. *Proceedings of 22nd International Conference Radioelektronika 2012, Brno, Czech Republic*, 1-4.
- [12]. Konuk, B., Zerman, E., Nur, G., & Akar, G. B. (2013). A spatiotemporal no-reference video quality assessment model. *2013 IEEE International Conference on Image Processing, Melbourne, Australia*, 54-58. Doi: 10.1109/ICIP.2013.6738012
- [13]. Bosse, S., Maniry, D., Müller, K.-R., Wiegand, T., & Samek, W. (2018). Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Transactions on Image Processing*, 27(1), 206-219. Doi: 10.1109/TIP.2017.2760518
- [14]. Naiwen, Y., & Donghui, Z. (2018). Applying the Convolutional Neural Network Deep Learning Technology to Behavioural Recognition in Intelligent Video. *Technical gazette*, 25(2), 528-535. Doi: 10.17559/TV-20171229024444
- [15]. Vlaović, J., Vranješ, M., Grabić, D., & Samardžija, D. (2019). Comparison of Objective Video Quality Assessment Methods on Videos with Different Spatial Resolutions. *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*, Osijek, Croatia, 287-292. Doi: 10.1109/IWSSIP.2019.8787324
- [16]. Rassool, R. (2017). VMAF reproducibility: Validating a perceptual practical video quality metric. *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Cagliari, Italy, 1-2. Doi: 10.1109/BMSB.2017.7986143
- [17]. Loke, M. H., Ong, E. P., Lin, W., Lu, Z., & Yao, S. (2006). Comparison of Video Quality Metrics on Multimedia Videos. *2006 International Conference on Image Processing, Atlanta, GA, USA*, 457-460. Doi: 10.1109/ICIP.2006.312492
- [18]. Zhang, H., Yuan, B., Dong, B., & Jiang, Z. (2018). No-Reference Blurred Image Quality Assessment by Structural Similarity Index. *Applied Sciences*, 10(8), 1-17. Doi: 10.3390/app8102003
- [19]. Lerga, J., Grbac, E., Sucic, V., & Saulig, N. (2018). Adaptive Methods for Video Denoising Based on the ICI, FICI, and RICl Algorithms. *Technical gazette*, 25, 1-6. Doi: 10.17559/TV-20130918123509
- [20]. Sara, U., Akter, M., & Uddin, M. (2019). Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. *Journal of Computer and Communications*, 7, 8-18.
- [21]. C. Brooks, A., Zhao, X., & N. Pappas, T. (2008). Structural Similarity Quality Metrics in a Coding Context: Exploring the Space of Realistic Distortions. *IEEE Transactions on Image Processing*, 17(8), 1261-1273. Doi: 10.1109/TIP.2008.926161
- [22]. Agarla, M., Celona, L., & Schettini, R. (2021). An Efficient Method for No-Reference Video Quality Assessment. *Journal of Imaging*, 7(3), 1-21. Doi: 10.3390/jimaging7030055
- [23]. Wang, Z., Simoncelli, E. P., & Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. *The Thirtieth Asilomar Conference on Signals, Systems & Computers*, 2, Pacific Grove, CA, USA, 1398-1402. Doi: 10.1109/ACSSC.2003.1292216
- [24]. Sedano, I., P riecto, G., Brunnström, K., Kihl, M., & Montalban, J. (2017). Application of full-reference video quality metrics in IPTV. In *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 1-4. IEEE.
- [25]. Qin, L., & Kang, L. (2018). Application of Video Scene Semantic Recognition Technology in Smart Video. *Technical gazette*, 25(5), 1429-1436. Doi: 10.17559/TV-20180620082101
- [26]. Bógalo, J., Poncela, P., & Senra, E. (2021). Circulant singular spectrum analysis: A new automated procedure for signal extraction. *Signal Processing*, 179, 1-17. Doi: 10.1016/j.sigpro.2020.107824

- [27]. Lu, Y., Kumar, J., Collier, N., Krishna, B., & Langston, M. A. (2018). Detecting Outliers in Streaming Time Series Data from ARM Distributed Sensors. *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, Singapore, 779-786. Doi: 10.1109/ICDMW.2018.00117
- [28]. Ford, W. (2015). The Algebraic Eigenvalue Problem - Chapter 18. In W. Ford, *Numerical Linear Algebra with Applications*, Academic Press, 379-438. Doi: 10.1016/B978-0-12-394435-1.00018-1
- [29]. Przystupa, K., Beshley, M., Hordiichuk-Bublivska, O., Kyryk, M., Beshley, H., Pyrih, J., & Selech, J. (2021). Distributed Singular Value Decomposition Method for Fast Data Processing in Recommendation Systems. *Energies*, 14(8), 1-24. Doi: 10.3390/en14082284
- [30]. Goran, N. (2021). *GitHub: gnermin/Elmar*. Retrieved from: <https://github.com/gnermin/Elmar> [accessed: 03 July 2023].